

**METHODS AND APPARATUS FOR SEMANTIC UNIT BASED AUTOMATIC  
INDEXING AND SEARCHING IN DATA ARCHIVE SYSTEMS**

**Abstract of the Disclosure**

An audio-based data indexing and retrieval system for processing audio-based data associated with a particular language, comprising: (i) memory for storing the audio-based data; (ii) a semantic unit based speech recognition system for generating a textual representation of the audio-based data, the textual representation being in the form of one or more semantic units corresponding to the audio-based data; (iii) an indexing and storage module, operatively coupled to the semantic unit based speech recognition system and the memory, for indexing the one or more semantic units and storing the one or more indexed semantic units; and (iv) a search engine, operatively coupled to the indexing and storage module and the memory, for searching the one or more indexed semantic units for a match with one or more semantic units associated with a user query, and for retrieving the stored audio based data based on the one or more indexed semantic units. The semantic unit may preferably be a syllable or morpheme. Further, the invention is particularly well suited for use with Asian and Slavic languages.

1500-66.APP